

音声認識技術を用いた歴史的音源のテキスト起こしに関する研究

Research on Text Transcription of Historical Sound Sources using Speech Recognition Technology

田中悠太

指導教員 三輪 賢一郎

サレジオ工業高等専門学校 機械電子工学科 情報コミュニケーション研究室

キーワード：音声認識, 文字起こし, SP レコード, API

1. 緒言

SP レコード(Standard Playing Record)は、1887 年から 1950 年代後半まで使用されたレコード及び録音技術である。全盛期には、流行歌、演説、講談など様々な音源の録音に使用されており、その歴史的価値は高い。

SP レコードのジャンルに演説、講談などの「語りもの」がある。それらは、SP レコード自体は現存しているものの、当時用いられた原稿の存在は不明である場合が少なくない。一部の落語などは当時の速記本が現存しているケースもあるが、必ずしも実際の録音と文言が完全一致するわけではないことが分かっている[1]。したがって、それら膨大なる文化遺産を文字の形で後世に残すためには、何らかの方法で文字起こしを実施する必要がある。しかしながら、このような文化遺産のアーカイブ化事業は、その性質上、国の予算で対応せざるを得ない。このことから、極力コストや労力のかからない方法が望ましい。

昨年度、本研究室では音声認識エンジンである Julius を使用して同様の研究を行ったが[2]、認識率が低く、現時点では使用に耐えうるものではないと判断した。そこで、本研究では比較的安価で使用でき、自らシステムを用意する必要のない API(Application Programming Interface)を用いた文字起こしの可能性を検証する。

2. 方法

本研究では AWS(Amazon Web Services)の Amazon Transcribe、Microsoft の azure の 2 つの音声認識 API を用いた。ハードウェアには、Windows OS 内蔵のノートパソコンを使用した。

また、API を使用する際のプログラムは Visual Studio Code を開発環境とし、Python を用いた。ノイズの除去には、Software 社の音声編集ソフトウェア WavePad を使用した。

今回対象とした音源は、「国立国会図書館デジタルコレクション「歴史的音源」」[3]に所蔵されている「ソロモン海戦に就いて 1」(日本コロムビア、昭和 19 年頃、収録時間 3 分 4 秒)を、国立国会図書館の許可のもとに使用した。

認識精度の評価指標としては、今回は仮名ベースでの認識パフォーマンスを考慮することとし、モーラ(拍)に着目した文字認識率を採用した。下記に文字認識率の算出式を示す。

文字認識率 =

$$\frac{\text{正解文字数} - \text{誤挿入文字数} - \text{誤削除文字数} - \text{誤置換文字数}}{\text{正解文字数}}$$

……(1)

3. 結果

上記のような方法で音声認識を実行した結果を図 1、図 2 に示す。

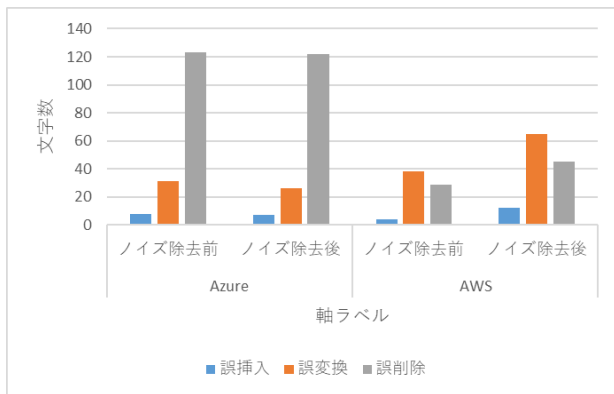


図1 各APIの認識結果

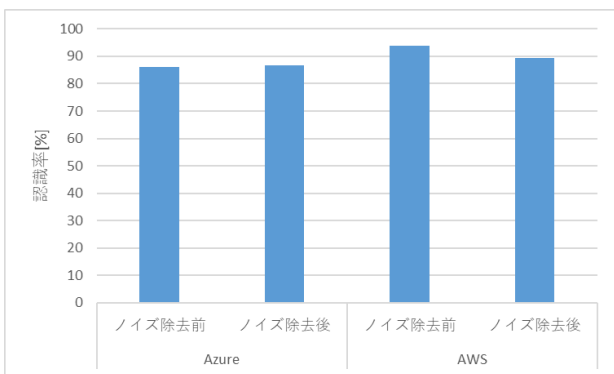


図2 各APIの認識率

図1は、各APIの認識結果とその内訳であるが、AWSに比べAzureは誤削除数が非常に多いことが分かる。これはAzureの認識において1文すべて認識できていない箇所が複数存在することが原因である。

図2からはAWSの認識率の方がAzureより若干高いこと、両方のAPIにおいてノイズ除去による認識率の向上は見られなかったことが分かる。この理由として、各APIの内部に既にノイズ除去の機能が備わっていることなどが考えられる。

4. 結 言

本研究では、音声認識APIを用いた歴史的音源の文字起こしの可能性について検証した。今回の結果から、どちらのAPIも86%~94%の高い認識率となっており、文字起こしとしての用途の可能性が確認された。しかし、Azureにおいては、1文すべて認識できないというような問題点の存在も明らかとなった。

また、今回の実験では、仮名をベースに認識率を計算したため、漢字の間違いを無視している。しかし、実際に文字起こしをする場合には、漢字の正答率は作業効率に直結する重要な指標である。ので、より実用的な指標も用いてAPIによる文字起こしの可能性を検証していく。

謝 辞

本研究は、国立国会図書館のご厚意により、「国立国会図書館デジタルコレクション 歴史的音源」[3]に所蔵の音源（同図書館ウェブサイトで公開中のもの）を用いております。

参考文献

- [1] 金澤裕之，“現代に繋がる近代初期の口語的資料における言語実態：速記本とSPレコードによる東西の落語を対象として，” 国立国語研究所論集，no. 10，pp. 58-84，2016年1月
- [2] 小泉朝陽，三輪賢一郎，“音声認識技術を用いた歴史的音源のテキスト起こしに関する研究，” 大学コンソーシアム八王子 第13回学生発表会，Q114，Dec. 2021.
- [3] 国立国会図書館デジタルコレクション「歴史的音源」Webサイト
(<https://rekion.dl.ndl.go.jp/>)